Critical Issues for Developing and Maintaining Item Banks and CATs: FDA Perspective

Robert O'Neill, PhD, Jane Scott, PhD, Laurie Burke, RPh, MPH

The Agency's limited experience with IRT, item banking and CAT prevents constructive feedback to researchers about use of these methods in medical product development programs. At a minimum, regulatory review issues will include those pertaining to any new effectiveness endpoint to measure and compare treatment-induced changes in clinical trials.

Generally, evidence of drug effectiveness is deemed substantial for claims in product labels or advertising if supported by adequate and well-controlled studies using endpoints that reliably and validly measure the specific concept(s) claimed. Important steps in the development and validation of PRO measurements as endpoints include the usual psychometric validation strategies such as the identification of the conceptual framework, item generation and reduction, and establishment of the measurement properties for the specific population targeted. IRT and CAT exploration in the clinical trial setting will assist FDA in determining how to apply established measurement principles to endpoints derived according to these methods.

The most critical consideration during FDA's review of the primary and secondary endpoints used by sponsors to measure treatment induced effects is the adequacy and availability of a clearly established concept of measurement, i.e., what is the measure measuring and is it meaningful in some sense as a treatment benefit? Examples of other clinical trial design issues include attention to controlling for bias, quantification of meaningful effect sizes, calculation of study sample size to demonstrate treatment effects of that size, and the relationships among all trial endpoints.

As with the assessment of clinical endpoints in any clinical trial, there are issues concerning the multiplicity of endpoints and controlling false positive conclusions, implications of missing data from subjects not completing assigned treatment or remaining in study, and potential heterogeneity of treatment effects among identified subgroups of study

subjects. It would seem appropriate to evaluate the performance of IRT strategies against other known PRO testing and measurement paradigms.

Tailoring items specific to an individual study subject introduces issues that are likely to involve considerable discussion by and interaction with the agency. It is too early for FDA to establish a position about whether or not item banking, IRT and CAT will be valuable in drug development. We encourage researchers to develop these tools in a few areas where they represent a distinct advantage over current measures. Discussions around these examples will provide the basis for informed FDA evaluations.